# Workflow for Hybrid Machine Learning Framework in Water Quality Assessment

## Step 1: Field Data Collection

Water samples are collected from 50 georeferenced sites across Bayelsa State, Nigeria. Locations are selected to represent urban, peri-urban, and rural settings. Samples are preserved and transported under standardized conditions.

## Step 2: Physicochemical Analysis

Ten key water quality parameters are measured in the laboratory: pH, EC, TDS, NO₃⁻, Cl⁻, SO₄²⁻, Ca²⁺, Mg²⁺, Na⁺, and Fe. Quality assurance and quality control (QA/QC) protocols are followed to ensure data accuracy.

## Step 3: WQI Computation

The Water Quality Index (WQI) is calculated using the weighted arithmetic mean method. Each parameter is normalized and weighted according to WHO standards, resulting in a single WQI score per location.

## Step 4: Machine Learning Modeling

Three ML models are trained for WQI prediction:  
- Random Forest (RF)  
- XGBoost  
- Long Short-Term Memory (LSTM)  
Models are evaluated on regression (R², MAE, RMSE) and classification (accuracy, F1-score, AUC).

## Step 5: Clustering Analysis

K-Means clustering is applied to the water quality dataset to uncover latent chemical groupings. Clusters are analyzed and mapped to interpret contamination signatures.

## Step 6: Spatial Visualization

Geographic Information System (GIS) tools and custom plots are used to visualize:  
- WQI Class Distribution  
- Model Performance  
- Spatial Clusters  
- Town-level Contamination Hotspots

## Step 7: Reporting & Policy Implication

Results are integrated into a hybrid decision-support system. Policy recommendations are made for environmental regulators, water utilities, and community stakeholders based on insights.